REMARKS

The examiner rejected claim 1 under 35 U.S.C. § 102(b) as being anticipated by Japanese Publication No. 2002/0031269 to Toshikazu Fukushima (Toshikazu). But contrary to what the examiner states, Toshikazu fails to disclose "in a large corpus, identifying geo-textual correlations among readings of the toponyms within the plurality of toponyms," as required by the claim. Instead, Toshikazu uses the presence of "co-occurring words" drawn from a look-up table to resolve an ambiguity in the meaning of a name. We explain the difference in more detail below.

Before we look at the meaning of "geo-textual correlations," it is important to understand what a reading of a toponym is. A reading of a toponym is a location with which the toponym is associated, such as a latitude-longitude location or an area. For example, a reading of "Paris" is the geographic region associated with Paris, France. Many toponyms have more than one reading; for example, another reading of Paris is the region associated with the town of Paris, Texas.

According to the instant specification, there is a relationship between a reading of a toponym and that toponym's location in a text: toponyms that have readings that are close to each other in space are more likely to be close to each other in a text. This is geo-textual correlation. The specification explains this further:

> A technical advance is achieved in the art by exploiting knowledge of a hitherto unobserved statistical property of documents, namely geo-textual correlation. By inspecting large corpora, we have found that there is a high degree of spatial correlation in geographic references that are in textual proximity. This applies not only to points that are nearby (such as Madison and Milwaukee), but also to geographic entities that enclose or are enclosed by regions (Madison and Wisconsin, for example). More specifically, if the textual distance between names N and M is small, and if N has a reading P (i.e., N is associated with P or N means P) and M has a reading Q, then the physical distance between P and Q is likely to be lower than would be expected randomly. Conversely, if P and Q are close geographically, then their names N and M are more likely to appear together in texts than would be expected randomly. This correlation between geographic and textual distance is considered in estimating of the confidence c(N,P) that a name N refers to a particular point P. (page 7, line 17-28, emphasis added)

Thus, for example, since Madison and Milwaukee are geographically close to each other, i.e., they have readings that are close to each other, the words "Madison" and "Milwaukee" are

statistically likely to appear close to each other within the text of a document in which they appear. Conversely, if the words Madison and Milwaukee often appear close to each other within a large corpus of documents, then they are statistically likely to have readings that are close to each other.

Claim 1 recites "in a large corpus, identifying geo-textual correlations among readings of ... toponyms." The examiner appears to believe that Toshikazu discloses this, and directs our attention to a two-paragraph passage. The first paragraph reads as follows:

> Incidentally, there are varieties of calculation methods in terms of appearance frequency information of co-occurring words in plural texts. In FIG. 9, for example, the location of "Chuo-ku" in the text 19 is not to be specified by referring to the referring link text 17, in which both co-occurring words "Tokyo" and "Osaka" appear. Consequently, according to the process (D), the analysis is performed referring to plural referring link texts. Additionally, even a linked text(s) is subject to the reference. Referring to the linked text 20 as well as the referring link texts 17 and 18, it is turned out that the co-occurring words "Tokyo", "Kinki-Area", and "Kyoto" appear in the texts once respectively, and "Osaka" appears three times. Thus "Chuo-ku" can be taken as "Chuo-ku" in Osaka in recognition of that "Osaka" makes the most of appearance. (¶ [0079], emphasis added)

But this paragraph does not disclose identifying geo-textual correlations in a large corpus of documents. The term "geo-textual correlations" implies a statistical analysis of a corpus of documents – typically a large corpus to make the statistical observations meaningful. Toshikazu does not perform any kind of statistical analysis of a corpus of documents to generate any correlations. Rather, Toshikazu's method involves resolving the ambiguity in the meaning of a location name by counting the frequencies of the appearance of co-occurring words, which he retrieves from a "named entity dictionary." (Fig. 7). The named entity dictionary is assembled by methods referred to in Toshikazu's background section, which include using pre-existing databases populated by various methods of extracting proper nouns. (See, e.g., ¶ [0008].) But none of these methods, nor anything mentioned in the above paragraph disclose performing a statistical analysis on a large corpus of documents to identify geo-textual correlations among readings of the toponyms, as required by the claim. Indeed, Toshikazu has no need for carrying out such statistical analyses because he already has the data he needs in the form of his named entity table, which he compiles from pre-existing databases.

The following is the second paragraph referred to by the examiner:

> In the above method, the co-occurring word that appears most frequently in the plural texts has priority. On the other hand, there is another method in which the co-occurring word that appears in the most numbers of referring link and linked texts has priority. Referring to FIG. 9, for example, the co-occurring word "Osaka" appears in the three texts, 17, 18, and 20, while "Kinki-Area" and "Kyoto" make their appearance in only the text 18. Thereby "Osaka" is regarded as the co-occurring word appearing in the most numbers of texts, and thereby used as a clue to resolve the ambiguity in the candidate named entity. (¶ [0080])

This paragraph discloses a close variant of Toshikazu's toponym ambiguity-resolving method that exploits the hyperlink structure of his texts. Here, the co-occurring word that appears in the greatest number of pages linked to the toponym to be resolved is selected to resolve the ambiguity. As in the previous paragraph, the method involves looking up a list of co-occurring words from a pre-existing table. And again, there is no reference to "in a large corpus, identifying geo-textual correlations among readings of the toponyms within the plurality of toponyms," as required by the claim. Furthermore, we are unable to find even a hint of such a reference anywhere within Toshikazu.

Toshikazu fails to anticipate claim 1 for another reason. The claim recites "using the identified geo-textual correlations to generate a value for a <u>confidence</u> that [a] selected toponym refers to a corresponding geographic location." (emphasis added) But nowhere in the cited paragraphs, nor within the named entity table, nor anywhere else in the reference is there any mention of confidence values that Toshikazu's co-occurring words refer to a particular place, as required by the claim.

The examiner also rejected claim 10 under 35 U.S.C. § 102(b) as being anticipated by Toshikazu. Contrary to what the examiner states, Toshikazu does not disclose:

> a document that includes a plurality of toponyms for which there is a corresponding plurality of (toponym,place) pairs, there being associated with each (toponym,place) pair of said plurality of (toponym,place) pairs a corresponding value for a <u>confidence</u> that the toponym of that (toponym,place) pair refers to the place of that (toponym,place) pair...

as required by the claim. (emphasis added). The examiner appears to believe that a similar passage to the one he pointed to for claim 1 discloses this. But, as discussed above, these paragraphs involve documents having a set of names for which Toshikazu looks up "co-

8

occurring words," in a table. Figure 7 and the following paragraph describes the structure of his table:

> The named entity dictionary 33 stores a dictionary for identifying the candidate named entities. FIG. 7 shows the structure of the named entity dictionary. As shown in FIG. 7, the named entity dictionary contains potential categories 41, such as "location name", "personal name", and "organization name", for each term of the named entityes (sic) 40. (It happens that the categories include non-named entity when the term can be a common noun.) And further, the dictionary stores a <u>co-occurring word list</u> 42 for each category. It is preferable that not only the co-occurring words but also their positional condition (for example, "collocating with the named entity", etc.) is added to the co-occurring word list 42. (¶ [0063], emphasis added)

In other words, the table includes a list of named entities (Fig. 7, 40), and for each named entity, a list of co-occurring words (Fig. 7, 42). But the dictionary contains no <u>(toponym,place) pairs,</u> nor does it include any <u>confidence values</u> that the co-occurring words refer to a particular place, as required by the claim. We were also unable to find any mention of the (toponym,place) structure or of confidence values anywhere else within Toshikazu.

Claim 10 also requires boosting the value of the confidence for a selected (toponym,place) pair. As discussed above, Toshikazu does not store confidence values in his named entity table, nor anywhere else. Therefore he has no confidence values to boost.

The examiner further rejected claim 18 under 35 U.S.C. § 102(b) as being anticipated by Toshikazu. Claim 18 requires "...identifying a plurality of (toponym,place) pairs that is associated with the selected document, and for each identified (toponym,place) pair, obtaining and using a value for a <u>confidence</u> that the toponym of the (toponym,place) pair refers to the place." (emphasis added) As discussed above for claim 10, Toshikazu makes no mention of (toponym,place) pairs, nor does he disclose obtaining and using confidence values that a toponym refers to a place.
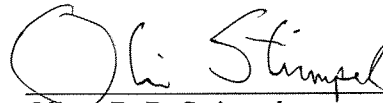
For the reasons discussed above, Applicant believes that claims 1, 10, and 18, and dependent claims 2-9 and 11-17 are not anticipated by Toshikazu.

In view of the above, Applicant believes the pending application is in condition for allowance.

Please charge the $525 fee for a three month extension of time, as well as any other fees
that might be due, or credit any overpayments to our Deposit Account No. 08-0219, under Order
No. 0113744.00124US2 from which the undersigned is authorized to draw.

Respectfully submitted,

Dated: December 19, 2007

Oliver B. R. Strimpel
Registration No.: 56,451
Attorney for Applicant

Wilmer Cutler Pickering Hale and Dorr LLP
60 State Street
Boston, Massachusetts 02109
(617) 526-6000 (telephone)
(617) 526-5000 (facsimile)

10